Gaetano Scaduto

Gli algoritmi di computer vision: una guida all'uso per le scienze politiche e sociali

(doi: 10.1424/104525)

Polis (ISSN 1120-9488)
Fascicolo 2, agosto 2022

Ente di afferenza:
Università degli studi di Milano Bicocca (unibicocca)

Gaetano Scaduto

Gli algoritmi di computer vision: una guida all'uso per le scienze politiche e sociali

Computer vision algorithms: user guide for social and political sciences

Automated techniques for image analysis represent a useful tool to broaden the frontiers of social research. The availability of commercial services relieves the researcher from the need to possess the technical and computational skills necessary to develop an algorithm autonomously, greatly expanding the audience of possible users. The following contribution aims to provide to the researchers the tools to understand the different characteristics of the main commercial services available as an instrument of research, while also offering a guide to use them and showing how previous studies have employed the new technique.

Keywords: Computer Vision; Machine Learning; Face Recognition; Emotion Recognition; Automated Image Analysis.

1. *Introduzione*

Negli ultimi anni gli algoritmi di machine learning hanno rivoluzionato le scienze sociali. Le tecniche di *quantitative text analysis* (QTA) hanno fornito ai ricercatori mezzi per l'analisi di testi dalle dimensioni inaccessibili alla codifica umana. Attraverso le stesse tecnologie, l'analisi delle immagini, prima dispendiosa in termini di tempo e denaro, ha oggi a disposizione strumenti in grado di aprire possibilità inedite: i servizi di computer vision. Si tratta di software basati su algoritmi di machine learning, in grado di ricevere in input un'immagine e restituire automaticamente in output una serie di informazioni, senza l'intervento umano.

La disponibilità di servizi commerciali per il riconoscimento automatizzato delle immagini solleva il ricercatore dalla necessità di possedere le capacità tecnico-informatiche e la potenza di calcolo necessarie a chi volesse cimentarsi in prima persona nello sviluppo di un algoritmo apposito. Le prestazioni dal punto di vista computazionale e la possibilità di standardizzare e riprodurre l'analisi, rendono l'utilizzo di questi servizi la scelta più immediata. Il crescente investimento in questo

campo e le sue innumerevoli applicazioni in ambiti come sicurezza, medicina e trasporti, rendono ottimisti sul continuo miglioramento delle prestazioni, dell'affidabilità e delle possibilità offerte.

Fra i principali utilizzi degli algoritmi di computer vision individuati dalla ricerca figurano: 1) rilevamento di volti (*face detection*); 2) riconoscimento facciale (*face recognition*), ovvero la capacità di riconoscere che due volti appartengono alla stessa persona; 3) classificazione, ovvero l'associazione di etichette (*tags*) inerenti al contenuto dell'immagine; 4) riconoscimento di attributi umani, ovvero di caratteristiche demografiche, stati emotivi e azioni compiute all'interno dell'immagine (Szeliski 2010; Joo e Steinert-Threlkeld 2018).

È possibile distinguere fra due classi di servizi di computer vision, sulla base del fatto che utilizzino algoritmi di machine-learning o meno. Il presente testo intende focalizzarsi sulla prima di queste due classi, in particolare sui servizi specializzati nel riconoscimento di attributi umani, estensivamente adoperati nella ricerca sociale[1].

È oggi disponibile una moltitudine di servizi commerciali di computer vision, ognuno con le proprie funzionalità e prezzi. Questo testo si concentrerà su quattro di essi: Face++, Microsoft Azure (Computer Vision API e Face API), Google Cloud Vision API e Kairos.

Nell'utilizzo dei servizi di computer, vanno anzitutto considerati aspetti che possono risultare eticamente problematici. Fra questi, spicca il tema della discriminazione algoritmica. Diversi studi (Raji *et al*. 2020; Serna *et al*. 2019) hanno svelato le distorsioni nella composizione demografica dei database su cui viene effettuato il *training* di questi algoritmi. Ciò ha l'effetto di rendere più accurati i risultati per le categorie maggiormente rappresentate (ad esempio: uomini caucasici) e meno accurati quelli per le altre categorie, specialmente quando le caratteristiche demografiche marginalizzate si sommano in una prospettiva intersezionale. È necessario avere contezza delle distorsioni prodotte dagli algoritmi, soprattutto per gli studi che intendano concentrarsi su un campione diversificato nel genere e nell'etnia. Va inoltre posta attenzione alla questione del rispetto della privacy nella raccolta delle immagini che si intenda analizzare (Jaeger *et al*. 2020) e sull'elevato impatto ambientale generato dal costo computazionale del *training* di un algoritmo di computer vision (Martineau 2020).

[1] Gli algoritmi non basati sul machine learning, sebbene meno accurati e flessibili, presentano vantaggi dal punto di vista computazionale, ad esempio nei compiti di *face detection*. Per un confronto fra queste due classi di algoritmi, si veda O'Mahony *et al*. (2019).

2. *Funzionamento e utilizzo di un software di computer vision basato su machine learning*

Una spiegazione intuitiva del funzionamento di questi servizi viene offerta da Mancosu e Bobba (2019). Gli studiosi illustrano come questi si basino su una famiglia di algoritmi di machine learning detta «reti neurali convoluzionali». (Convolutional neural network o CNN). Si tratta di modelli relativamente facili da sottoporre a *training* rispetto ad altre classi di reti neurali (Krizhevsky *et al*. 2012). L'algoritmo viene sottoposto ad un *training* con l'obiettivo di fornire ad esso un numero sufficiente di casi per poter riconoscere una caratteristica presente all'interno dell'immagine. Ogni immagine viene rappresentata come una matrice di pixel di diverse colorazioni: il colore di ogni pixel può essere espresso come la combinazione di tre valori rappresentanti la componente di rosso, verde e blu nella forma (R, G, B). Si scompone l'immagine in input in tre matrici di uguale dimensione contenenti, per ogni pixel, il numero corrispondente alla tonalità di rosso, verde e blu. Successivamente viene effettuata la procedura di convoluzione: una matrice di *feature detection* scorre all'interno dei livelli dell'immagine per trovare degli elementi salienti e produrre una serie di immagini convolute, all'interno delle quali sono evidenziate le principali caratteristiche. A seguito di ciò, le immagini convolute attraversano una procedura di «pooling», una serie di step attraverso cui la dimensionalità e la complessità dell'immagine vengono ridotte. Una volta definita una regione spaziale (n x n pixel) all'interno dell'immagine, l'algoritmo produce una mappa rettificata delle caratteristiche, che emerge effettuando una media dei pixel di quella regione riassunta in un singolo pixel. Il risultato è un vettore che contiene tutte le caratteristiche salienti dell'immagine originale. A questo punto, l'algoritmo confronta il vettore con il proprio *training set*, restituendo in output il grado di fiducia che, all'interno dell'immagine, la caratteristica oggetto di indagine sia presente.

Va notato come il riconoscimento di specifici attributi relativi all'immagine possa essere usato non solo come fine in sé, ma anche come strumento di riconoscimento facciale (Kumar *et al*. 2011). Non va sottovalutata l'onerosità computazionale del *training* di questi algoritmi, che necessita di un'elevata potenza di calcolo. Ciò ha portato a riconoscere la convenienza della dislocazione delle operazioni attraverso software che mettono a disposizione API (Application Programming Interfaces) commerciali, che permettono di effettuare queste operazioni presso server terzi (largamente più potenti dei normali personal compu-

ter), ricevendo in input l'immagine e restituendo in output le caratteristiche desiderate.

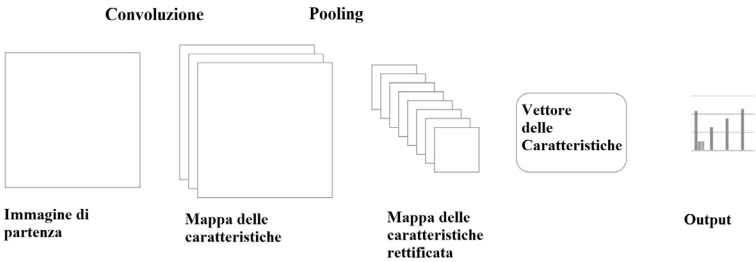

Fig. 1. Descrizione sintetica del funzionamento di un algoritmo di CV basato sul machine learning (Mancosu e Bobba 2019).

Il ricercatore che intendesse fare uso di API commerciali dovrà seguire, per ottenere i dati desiderati, una serie di passaggi.

Per accedere ai servizi è necessaria una registrazione. La maggioranza dei servizi offre dei periodi di prova gratuiti, dopo i quali è possibile scegliere fra diverse formule a pagamento, con limitazioni e sovrapprezzi in base al consumo. Per interagire con questi servizi, è necessario dialogare con i server attraverso un linguaggio di programmazione (come R o Python). È necessario fare uso di apposite librerie per effettuare le «chiamate» all'API. Fra le librerie utili al ricercatore possiamo annoverare: per Face++ *facepplib* (Python), per Microsoft Azure *AzureVision* (R, utilizzabile sia per Face API che per Computer Vision API), *azure-cognitiveservices-vision-computervision* (Python, utilizzabile per Computer Vision API) e *azure-cognitiveservices-vision-face* (Python, utilizzabile per Face API), per Google Cloud Vision *RoogleVision* o *googleCloudVisionR* (R) e *vision* dal pacchetto *google.cloud* (Python), per Kairos *facerec* (R) o *kairos_face* (Python).

Una volta presentate le chiavi d'accesso ed installate le librerie, è possibile inviare le immagini ai server. Queste possono essere immagini presenti all'interno del terminale o URL pubblicamente accessibili. Contestualmente, sarà necessario specificare gli attributi desiderati

```
{
 "request_id": "1652341181,a883f63e-eb94-4b29-98d8-db02208fcb3e",
 "time_used": 173,
 "faces": [
  {
   "face_token": "1d82f39e7d582e10dccd2c09304cb4f8",
   "face_rectangle": {
    "top": 187,
    "left": 942,
    "width": 101,
    "height": 101
   },
   "attributes": {
    "emotion": {
     "anger": 0,
     "disgust": 0.002,
     "fear": 0,
     "happiness": 99.994,
     "neutral": 0.003,
     "sadness": 0,
     "surprise": 0
    }
   }
  },
  {
   {
    "face_token": "aa58142c6056013bdaeb4cba7ae1f3a7",
    "face_rectangle": {
     "top": 188,
     "left": 145,
     "width": 94,
     "height": 94
    },
    "attributes": {
     "emotion": {
      "anger": 79.24,
      "disgust": 0.101,
      "fear": 0.101,
      "happiness": 0.595,
      "neutral": 5.663,
      "sadness": 10.678,
      "surprise": 3.622
     }
    }
   }
  ],
  "image_id": "K8LO6Jw880O52w18TBfSyA==",
  "face_num": 2
 }
```


Fig. 2. Output in formato JSON fornito da Face++.

come output. Nel caso la chiamata sia effettuata con successo, l'output consiste tipicamente in un oggetto JSON (JavaScript Object Notation). Si tratta di un formato che sarà necessario trasformare in un dataset attraverso operazioni di *parsing* (Jaeger *et al*. 2020). Un esempio di output in formato JSON proveniente da Face++, relativo all'immagine in figura 4, è riportato di seguito.

Al fine di fornire al ricercatore gli elementi necessari ad effettuare una decisione consapevole sui servizi da utilizzare, saranno adesso presentate le principali funzionalità di quattro fra i servizi commerciali più utilizzati: Face++, Microsoft Azure, Google Cloud Vision e Kairos.

Face++, attraverso la Detect API, fornisce la possibilità di ottenere informazioni sui volti presenti all'interno dell'immagine, in particolare: sesso, età, intensità del sorriso, etnia, presenza di occhiali, maschere e bagaglio emotivo presente nella foto. Quest'ultimo riguarda le emozioni espresse dai volti secondo la percezione umana e viene articolato in: rabbia, disgusto, paura, felicità, neutralità, tristezza e sorpresa. Ad ogni emozione viene associato un valore da 0 a 100 indicante il grado di fiducia sulla presenza dell'emozione nel volto esaminato.

Microsoft Azure mette a disposizione due diverse API per l'analisi delle immagini. La prima è la Computer Vision API che, data un'immagine in input, restituisce una stringa di testo contenente una descrizione del contenuto dell'immagine. Questa descrizione è composta dal soggetto (o soggetti) e dall'azione compiuta. Il soggetto può essere un termine generico, collettivo o un nome proprio di persona, qualora questa sia un personaggio pubblico che la macchina virtuale è in grado di riconoscere.

La seconda API di Microsoft Azure è la Face API. Essa è in grado di restituire età, sesso, intensità del sorriso (un valore compreso fra 0 e 1), presenza o meno di capelli, occhiali, make-up, accessori (cappelli o maschere) e il grado di fiducia riguardo alla presenza di rabbia, disprezzo, disgusto, paura, felicità, neutralità, tristezza e sorpresa.

Google Cloud Vision presenta una serie di funzionalità diverse rispetto ai concorrenti. È in grado di restituire la presenza di gioia, tristezza, rabbia e sorpresa. Oltre a ciò, restituisce il nome degli oggetti presenti nella foto, anche direttamente sul volto della persona e una stringa di etichette. Una funzionalità offerta esclusivamente da questo servizio è quella di rilevare la presenza di contenuti espliciti (*safe search*). Questi sono organizzati in cinque categorie: *adult* (nudità e pornografia), *spoof* (manipolazioni nell'immagine), *medical*, *violence* e *racy* (contenuti ammiccanti non esplicitamente pornografici). Il grado di fiducia riguardo le emozioni e gli attributi di *safe search* è espressa attraverso una scala a cinque livelli (da «molto improbabile» a «molto probabile»), mentre quello riguardante gli oggetti e le etichette attraverso una percentuale.

Kairos è in grado di restituire caratteristiche demografiche come età, etnia e sesso, oltre all'eventuale presenza di occhiali ed informazio-

ni riguardo gioia, sorpresa, disgusto, rabbia, paura e tristezza espresse dai volti nell'immagine. Tutte le informazioni sono accompagnate dal grado di fiducia, espresso con un valore compreso fra zero e uno.

	Face++	Microsoft Azure	Google Cloud Vision	Kairos
Riconoscimento Facciale	✓	✓	×	✓
Età	✓	✓	×	✓
Sesso	✓	✓	×	✓
Etnia	✓	×	×	✓
Occhiali	✓	✓	✓	✓
Makeup	×	✓	×	×
Accessori	✓	✓	✓	×
Mascherine	✓	✓	×	×
Felicità	✓	✓	✓	✓
Disprezzo	×	✓	×	×
Disgusto	✓	✓	×	✓
Rabbia	✓	✓	✓	✓
Sorpresa	✓	✓	✓	✓
Paura	✓	✓	×	✓
Neutralità	✓	✓	×	×
Tristezza	✓	✓	✓	✓
Contenuti espliciti	×	×	✓	×
Etichette	×	✓	✓	×
Descrizione del contenuto	×	✓	×	×

Fig. 3. Funzionalità supportate dai software di computer vision.

Alcune informazioni rilevate da questi servizi possono risultare problematiche dal punto di vista etico. Oltre alle già citate questioni nel rilevamento di etnia e sesso, alcune misure, come il punteggio di bellezza reso disponibile da Face++, possono essere influenzate da criteri di arbitrarietà. Va anche notata la scelta, da parte di Google, di non includere il riconoscimento del sesso per ragioni etiche (Ghosh 2020).

3. *Performance e affidabilità*

Nella scelta di un software di computer vision vanno considerate le differenze, in termini di qualità e performance, fra i diversi servizi. A tal fine, verranno di seguito presentati dati riguardanti prestazioni e affidabilità dei quattro servizi esposti in precedenza. Si noti che gli studi riportati di seguito testano l'accuratezza dei servizi su database differenti.

Ciò porta a suggerire cautela nel confronto, per via della distorsione nei risultati introdotta dall'utilizzo di diversi set di immagini.

Riguardo le caratteristiche demografiche, Dehghan e colleghi (2017) osservano come l'errore medio assoluto nella stima dell'età di Microsoft Azure sia 7,62 anni, inferiore a quella di Kairos (10,57 anni) e Face++ (11,04 anni). Per il sesso, Microsoft avrebbe un'accuratezza del 90,86%, Kairos dell'84,66%, Face++ dell'83,04%. Dati confermati da Mancosu e Bobba (2019), che osservano uno scarto quadratico medio di 6,3 anni per l'età e un'accuratezza del 97% per il sesso. Jung e colleghi (2018) rilevano come Face++ e Microsoft Azure assegnino correttamente il sesso rispettivamente nel 92% e 97% dei casi, mentre l'età viene correttamente assegnata nel 34% e 45% dei casi, con Azure che produrrebbe complessivamente errori meno pronunciati. Studi recenti confermano l'affidabilità di Microsoft Azure nei compiti di *face detection* e nell'individuazione di sesso ed età, con performance che non si discostano eccessivamente da quelle di Kairos (Malone e Burns 2021). Bakhshi e colleghi (2014) rivelano un accordo fra le informazioni fornite da Face++ e quelle ricavate da coder umani del 97% per la *face detection*, del 96% per il riconoscimento del sesso (si veda anche Chakraborty *et al*. 2017) e fra il 93% ed il 99% per l'appartenenza ad una fascia d'età. Più conservative le stime di Chakraborty e colleghi (2017), che rivelano un accordo con la codifica umana del 88% per il sesso, del 79% per l'etnia e dell'83% per l'età. La stessa Face++ (Zhou *et al*. 2015) riporta un'accuratezza, per la *face recognition* del 99,5%.

Per quanto riguarda il riconoscimento delle emozioni, Peng (2018), analizzando Google Cloud Vision, Face++ e Microsoft Azure, rileva come quest'ultimo mostri una correlazione più alta dei concorrenti rispetto alla codifica umana per ogni emozione presa in esame, sebbene Face++ sia più performante nei compiti di *face detection* (Peng 2018; Jung *et al*. 2018). Altre evidenze (Khanal *et al*. 2018) suggeriscono che Microsoft performi meglio di Google nel riconoscimento delle emozioni, individuando un 60% di veri positivi contro il 45,25% di Google. Dehghan e colleghi (2017) assegnano ai servizi di Microsoft un'accuratezza del 61,3% in questo campo. Va notato come questa accuratezza sia variabile fra le diverse emozioni, risultando affidabile soprattutto per le manifestazioni di felicità e neutralità (Peng 2018). Deshmukh e colleghi (2017) classificano come alta l'accuratezza di Microsoft Azure nel riconoscere emozioni e come moderata quella di Kairos. Il giudizio di Kairos riguardo la presenza di una determinata emozione si discoste-

rebbe da quello degli annotatori umani nel 40% dei casi (Dupré *et al*. 2018)[2].

I compiti di associazione delle etichette alle immagini comportano una serie di complicazioni dal punto di vista computazionale e metodologico. Ciò rende difficile la valutazione delle performance per queste operazioni. Degno di menzione, in questo campo, risulta essere il lavoro di Chen e Chen (2017), che hanno riconosciuto a Google Cloud Vision un'accuratezza media del 42,4% nel riconoscimento delle etichette. Altre evidenze suggeriscono un'affidabilità di questo software superiore del 15% rispetto alla codifica umana, che fornirebbe descrizioni meno dettagliate in termini di numero di attributi e incapperebbe maggiormente in interpretazioni erronee del contenuto (O'Neal *et al*. 2018).

Fig. 4. L'immagine sottoposta all'analisi dei servizi (fonte: Wikimedia Commons).

[2] Dupré e colleghi dicotomizzano l'output restituito dai servizi di computer vision considerando l'emozione come riconosciuta se il grado di fiducia è superiore a 0,5, non riconosciuta altrimenti. Altri studi (Khanal *et al*. 2018), considerano invece come riconosciuta solo l'emozione per la quale viene restituito il grado di fiducia maggiore.

Va posta attenzione a come i software reagiscono alla distorsione nella presentazione dei volti. Microsoft Azure non sembrerebbe in grado di assolvere pienamente ai compiti di *face detection* quando un viso è presentato di profilo, mentre Google Cloud Vision non soffre di questa mancanza (Khanal *et al*. 2018). Inoltre, luminosità, prospettiva, dimensioni e oscillazioni possono influenzare pesantemente le prestazioni (Malone e Burns 2021). Un aspetto che merita ulteriori approfondimenti, alla luce degli sviluppi recenti, è quello relativo alla capacità di riconoscere i volti parzialmente coperti da mascherine (Fulzele *et al*. 2021).

A fini illustrativi, viene di seguito riportata l'analisi di un'immagine attraverso Face++ e Google Cloud Vision. L'immagine sottoposta ai servizi raffigura Sergio Mattarella e Giuseppe Conte nel 2018.

Tab 1. *I risultati forniti da Face++*

	Faccia 1 (Mattarella)	Faccia 2 (Conte)
Felicità	0,63%	99,99%
Neutralità	10,11%	0,01%
Sorpresa	43,20%	0%
Tristezza	0,77%	0%
Disgusto	0,14%	0%
Rabbia	45,02%	0%
Paura	0,14%	0%
Età	85	47
Genere	Maschio	Maschio
Sorriso	35,93%	99,93%
Occhio sinistro	Aperto, Occhiali	Aperto, No occhiali
Occhio destro	Aperto, Occhiali	Aperto, No occhiali

I risultati riportati nelle tabelle precedenti rappresentano solo una parte delle funzionalità offerte dai servizi. È però possibile formulare alcune osservazioni. I due servizi discordano sulla lettura del bagaglio emotivo di Sergio Mattarella. Inoltre, Face++ non assegna l'età corretta, ma l'errore non si discosta eccessivamente dal valore reale (nell'immagine Conte ha 53 anni, Mattarella 76), mentre fornisce informazioni corrette su genere e presenza di occhiali. Per quanto riguarda il rilevamento di oggetti ed etichette di Cloud Vision, esso sembra soddisfacente, nonostante la discordanza fra il grado di fiducia assegnato

a «flower» come etichetta e come oggetto. Va sottolineato come le inesattezze prodotte nell'analisi delle singole immagini possano perdere di rilevanza qualora si conducano analisi su un ampio corpus di immagini, almeno nel caso in cui siano assenti distorsioni sistematiche negli algoritmi adoperati.

Tab. 2. *I risultati forniti da Google Cloud Vision*

Faccia 1 (Mattarella)		Faccia 2 (Conte)	
Gioia	Probabile	Gioia	Molto probabile
Tristezza	Molto improbabile	Tristezza	Molto improbabile
Rabbia	Molto improbabile	Rabbia	Molto improbabile
Sorpresa	Molto improbabile	Sorpresa	Molto improbabile
Esposto	Molto improbabile	Esposto	Molto improbabile
Sfocato	Molto improbabile	Sfocato	Molto improbabile
Cappelli	Molto improbabile	Cappelli	Molto improbabile
Fiducia (%)	94	Fiducia	97

Oggetti	Fiducia (%)	Etichette	Fiducia (%)	Ricerca sicura (%)	
Cravatta	94	Fotografia	94	Adult	Molto improbabile
Cravatta	92	Cornice	93	Spoof	Molto improbabile
Persona	80	Sorriso	91	Medical	Molto improbabile
Persona	79	Fiore	90	Violence	Molto improbabile
Cornice	73	Cravatta	89	Racy	Molto improbabile
Persona	71	Cappotto	89		
Persona	70	Pianta	84		
Fiore	64	Abito	84		

4. *L'utilizzo della computer vision nelle scienze politiche e sociali*

Quali sono state le principali applicazioni degli algoritmi di computer vision nella ricerca? Per rispondere a questa domanda, verrà presentata una serie di studi che hanno utilizzato questi algoritmi, sia in forma autoprodotta che attraverso i servizi commerciali sopra esposti, nelle scienze politiche e sociali.

Nella comunicazione politica, fondamentale risulta essere il lavoro di Joo e colleghi (2014), dove viene creato un modello che fa uso di una serie di attributi dell'immagine per ottenere informazioni sugli intenti comunicativi. Il modello, data un'immagine in input, estrapola degli indicatori elementari (sorriso, saluto, gesti, ecc.) e li utilizza per inferire la *favorability* dell'immagine. Questa tecnica, applicata alle fotografie

raffiguranti Barack Obama, ha rivelato importanti correlazioni fra la *favorability* assegnata dall'algoritmo ed il tasso di approvazione rilevato dai sondaggi. Lo stesso gruppo di ricercatori (Joo *et al*. 2015) ha messo a punto un modello in grado di inferire, dalle fotografie dei candidati, specifici tratti sociali (intelligenza, onestà, ecc.), dai quali viene costruito un modello in grado di predire l'esito di elezioni senatoriali e governatoriali negli USA con un'accuratezza superiore al 65%.

Recenti studi hanno utilizzato il servizio proprietario di Microsoft per indagare gli effetti delle manifestazioni emotive sulla percezione dei candidati da parte dell'elettorato durante un dibattito televisivo, sia nel contesto statunitense (Boussalis e Coan 2021), che tedesco (Boussalis *et al*. 2021). L'utilizzo del medesimo software ha contribuito anche a svelare le differenze quantitative nelle manifestazioni di felicità fra politici populisti e mainstream, e come queste influenzino i risultati elettorali (Masch *et al*. 2021). Peng (2018) ha mostrato, attraverso l'utilizzo di Face++ e Microsoft Azure, le modalità attraverso cui i *bias* dei media influenzano la selezione delle immagini dei candidati avversi, esaltandone le caratteristiche negative.

I servizi di computer vision si sono rivelati particolarmente utili per le ricerche condotte nel campo dell'analisi della comunicazione via social media. Zhu e colleghi (2013) hanno utilizzato questi algoritmi per ricavare informazioni sull'engagement generato dalle immagini a contenuto politico su Flickr. Queste sono state analizzate attraverso un algoritmo in grado di rilevare la presenza del volto del candidato, di testo e del logo della campagna elettorale. Questi elementi risultavano legati ad un aumento di visualizzazioni e commenti. Una ricerca successiva (Bakhshi *et al*. 2014), analizzando oltre un milione di immagini provenienti da Instagram attraverso Face++, confermerà la correlazione fra presenza di volti e aumento di like e commenti. Una correlazione rilevata anche in ambito politico da Peng (2021) che, utilizzando Face++ sulle fotografie pubblicate dai politici americani su Instagram, osserva come la presenza di volti ed il loro contenuto emozionale siano legati ad un aumento del numero di like, soprattutto se questi volti sono quelli dei politici stessi, ed in maniera proporzionale all'intensità dell'emozionalizzazione e all'area occupata dal volto nell'immagine.

I dati demografici ricavati dalle immagini del profilo di Facebook e Twitter, estratti attraverso Microsoft Azure e Face++, hanno permesso di ottenere informazioni riguardo l'età dei follower di Trump e Clinton durante la campagna per le elezioni presidenziali del 2016 (Wang *et al*. 2016) e si sono rivelati in accordo con le informazioni ri-

levate dai sondaggi nel contesto del referendum sulla Brexit (Mancosu e Bobba 2019). Face++ è anche stato utilizzato per svelare la composizione demografica degli utilizzatori di determinati hashtag su Twitter (Chakraborty *et al*. 2017) e per classificare, attraverso età e genere, il contenuto dei selfie pubblicati su Instagram (Deeb-Swihart *et al*. 2017).

Gli studi che hanno utilizzato Google Cloud Vision hanno sfruttato soprattutto la funzionalità meglio sviluppata di questo servizio: l'associazione delle etichette. Queste sono state utilizzate per costruire dei predittori di tratti della personalità degli utenti a partire dalle immagini pubblicate su Instagram (Ferwerda e Tkalcic 2018), per svelare gli intenti comunicativi nascosti dietro un corpus di 5000 immagini su Twitter (Razis *et al*. 2021) e per l'analisi di milioni di immagini prodotte dai *troll* russi durante le elezioni americane del 2016 (Zannettou *et al*. 2020).

Kairos risulta il meno diffuso fra i servizi presi in considerazione. Va segnalato il suo utilizzo in una ricerca finalizzata a svelare la composizione degli utenti di Twitter in Romania in termini di età, etnia e genere (Florea e Roman 2018).

5. *Conclusioni*

L'utilizzo degli algoritmi di computer vision per la ricerca sociale ha prodotto finora risultati incoraggianti. Ciononostante, i margini di miglioramento sono ampi, sia dal punto di vista dell'accuratezza che delle possibilità offerte[3]. La crescita delle prestazioni è destinata ad ampliare lo spettro delle discipline che faranno uso di questi algoritmi: dal marketing alla comunicazione pubblica, fino all'antropologia e agli studi culturali, ad esempio in campo artistico. È anche immaginabile che nel prossimo futuro si presenterà la possibilità di utilizzare, in modo altrettanto immediato, software analoghi per l'analisi dei video (già presenti in alcuni dei servizi descritti nel presente testo). Ciò aprirebbe una serie di opportunità inedite, permettendo l'analisi di contenuti televisivi e cinematografici, dai canali «all-news» alle piattaforme di streaming. Inoltre, l'analisi dei contenuti pubblicati sui social, che sempre di più

[3] Per quanto popolare, l'accuratezza è solo uno dei diversi aspetti da prendere in considerazione per considerare la validità di un servizio di computer vision. Fra gli altri fattori vanno citati, ad esempio, la stabilità nei confronti del rumore e la consistenza logica delle analisi prodotte (si vedano Ribeiro *et al*. 2020; Geirhos *et al*. 2021).

investono sulla dimensione video (TikTok, Instagram Reels, ecc.), potrà essere condotta con strumenti adatti alle evoluzioni comunicative. Vi è dunque ragione di essere ottimisti sulla prospettiva futura delle ricerche nel campo dell'immagine.

Riferimenti bibliografici

Bakhshi, S., Shamma, D.A. e Gilbert, E. (2014). «Faces Engage Us: Photos with Faces Attract More Likes and Comments on Instagram». In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 965-974. New York: ACM.

Boussalis, C. e Coan, T.G. (2021). «Facing the Electorate: Computational Approaches to the Study of Nonverbal Communication and Voter Impression Formation». In *Political Communication*, 38 (1-2), pp. 75-97.

Boussalis, C., Coan, T.G., Holman, M.R. e Müller, S. (2021). «Gender, Candidate Emotional Expression, and Voter Reactions during Televised Debates». In *American Political Science Review*, 115 (4), pp. 1242-1257.

Chakraborty, A., Messias, J., Benevenuto, F., Ghosh, S., Ganguly, N. e Gummadi, K. (2017). «Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations». In *Proceedings of the International AAAI Conference on Web and Social Media*, pp. 22-31. Palo Alto, Calif,: The AAAI Press.

Chen, S.H. e Chen, Y.H. (2017). «A Content-Based Image Retrieval Method Based on the Google Cloud Vision API and Wordnet». In *Intelligent Information and Database Systems. ACIIDS 2017: Lecture Notes in Computer Science*, a cura di N. Nguyen, S. Tojo, L. Nguyen e B. Trawiński, B, pp. 651-662. Cham: Springer.

Deeb-Swihart, J., Polack, C., Gilbert, E. e Essa, I. (2017). «Selfie-Presentation in Everyday Life: A Large-Scale Characterization of Selfie Contexts on Instagram». In *Proceedings of the International AAAI Conference on Web and Social Media*, pp. 42-51. Palo Alto, Calif,: The AAAI Press.

Dehghan, A., Ortiz, E.G., Shu, G. e Masood, S.Z. (2017). «Dager: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network». ArXiv:1702.04280.

Deshmukh, R. S. e Jagtap, V. (2017). «A Survey: Software Api and Database for Emotion Recognition». In *2017 International Conference*

on Intelligent Computing and Control Systems (ICICCS), pp. 284-289. Piscataway, N.J.: IEEE.

Dupré, D., Andelic, N., Morrison, G. e McKeown, G. (2018). «Accuracy of Three Commercial Automatic Emotion Recognition Systems across Different Individuals and Their Facial Expressions». In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops*, pp. 627-632. Piscataway, New York: IEEE.

Ferwerda, B. e Tkalcic, M. (2018). «Predicting Users' Personality from Instagram Pictures: Using Visual and/or Content Features?». In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, pp. 157-161. New York: ACM.

Florea, A. e Roman, M. (2018). «Using Face Recognition with Twitter Data for the Study of International Migration». In *Informatica Economica*, 22 (4), pp. 31-46.

Fulzele, V., Kirad, P., Dubey, C., Kulkarni, Y. e Thatte, S. (2021). «Utilizing Cloud Capabilities for Face Detection and Face Recognition During COVID-19: Comparative Analysis». In *Proceedings of the International Conference on Smart Data Intelligence*, pp. 1-12.

Geirhos, R., Meding, K. e Wichmann, F.A. (2020). «Beyond Accuracy: Quantifying Trial-by-Trial Behaviour of CNNs and Humans by Measuring Error Consistency». In *Advances in Neural Information Processing Systems*, 33, pp. 13890-13902.

Ghosh, S. (2020). «Google AI Will No Longer Use Gender Labels Like "Woman" or "Man" on Images of People to Avoid Bias». In *Business Insider Nederland*, 20 febbraio.

Jaeger, B., Sleegers, W.W. e Evans, A.M. (2020). «Automated Classification of Demographics from Face Images: A Tutorial and Validation». In *Social and Personality Psychology Compass*, 14 (3), e12520.

Joo, J. e Steinert-Threlkeld, Z.C. (2018). «Image as Data: Automated Visual Content Analysis for Political Science». ArXiv:1810.01544.

Joo, J., Li, W., Steen, F.F. e Zhu, S.C. (2014). «Visual Persuasion: Inferring Communicative Intents of Images». In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 216-223. Piscataway, N.J.: IEEE.

Joo, J., Steen, F.F. e Zhu, S.C. (2015). «Automated Facial Trait Judgment and Election Outcome Prediction: Social Dimensions of Face». In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3712-3720. Piscataway, N.J.: IEEE.

Jung, S.G., An, J., Kwak, H., Salminen, J. e Jansen, B.J. (2018). «Assessing the Accuracy of Four Popular Face Recognition Tools for Inferring Gender, Age, and Race». In *Twelfth International AAAI Conference on Web and Social Media*, pp. 624-627. Palo Alto, Calif: The AAAI Press.

Khanal, S.R., Barroso, J., Lopes, N., Sampaio, J. e Filipe, V. (2018). «Performance Analysis of Microsoft's and Google's Emotion Recognition API Using Pose-Invariant Faces». In *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion*, pp. 172-178. New York: ACM.

Krizhevsky, A., Sutskever, I. e Hinton, G.E. (2012). «ImageNet Classification with Deep Convolutional Neural Networks». In *Advances in Neural Information Processing Systems*, 25, pp. 1097-1105.

Kumar, N., Berg, A., Belhumeur, P.N. e Nayar, S. (2011). «Describable Visual Attributes for Face Verification and Image Search». In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33 (10), pp. 1962-1977.

Malone, A. e Burns, J. (2021). «Evaluating the Accuracy of Public Cloud Vendor Face Detection API's». In *Journal of Image and Graphics*, 9 (1), pp. 20-26.

Mancosu, M. e Bobba, G. (2019). «Using Deep-Learning Algorithms to Derive Basic Characteristics of Social Media Users: The Brexit Campaign as a Case Study». In *PloS One*, 14 (1), e0211013.

Martineau, K. (2020). «Shrinking Deep Learning's Carbon Footprint». In *The MIT News*, 7 agosto.

Masch, L., Gassner, A. e Rosar, U. (2021). «Can a Beautiful Smile Win the Vote?: The Role of Candidates' Physical Attractiveness and Facial Expressions in Elections». In *Politics and the Life Sciences*, 40 (2), pp. 213-223.

O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G.V., Krpalkova, L. e Walsh, J. (2019). «Deep Learning vs. Traditional Computer Vision». In *Science and Information Conference*, pp. 128-144. New York: Springer.

O'Neal, A., Rodgers, B., Segler, J., Murthy, D., Lakuduva, N., Johnson, M. e Stephens, K. (2018). «Training an Emergency-Response Image Classifier on Signal Data». In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 751-756. Piscataway: N.J. IEEE.

Peng, Y. (2018). «Same Candidates, Different Faces: Uncovering Media Bias in Visual Portrayals of Presidential Candidates with Computer Vision». In *Journal of Communication*, 68 (5), pp. 920-941.

— (2021). «What Makes Politicians' Instagram Posts Popular? Analyzing Social Media Strategies of Candidates and Office Holders with Computer Vision». In *The International Journal of Press/Politics*, 26 (1), pp. 143-166.

Raji, I.D., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J. e Denton, E. (2020). «Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing». In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 145-151. New York: ACM

Razis, G., Theofilou, G. e Anagnostopoulos, I. (2021). «Latent Twitter Image Information for Social Analytics». In *Information*, 12 (2), pp. 49-72.

Ribeiro, M.T., Wu, T., Guestrin, C. e Singh, S. (2020). «Beyond Accuracy: Behavioral Testing of NLP Models with CheckList». ArXiv:2005.04118.

Serna, I., Morales, A., Fierrez, J., Cebrian, M., Obradovich, N. e Rahwan, I. (2019). «Algorithmic Discrimination: Formulation and Exploration in Deep Learning-Based Face Biometrics». ArXiv:1912.01842.

Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. New York: Springer.

Wang, Y., Li, Y. e Luo, J. (2016). «Deciphering the 2016 Us Presidential Campaign in the Twitter Sphere: A Comparison of the Trumpists and Clintonists». In *Tenth International AAAI Conference on Web and Social Media*, pp. 723-726. Palo Alto, Calif.: The AAAI Press.

Zannettou, S., Caulfield, T., Bradlyn, B., De Cristofaro, E., Stringhini, G. e Blackburn, J. (2020). «Characterizing the Use of Images in State-Sponsored Information Warfare Operations by Russian Trolls on Twitter». In *Proceedings of the International AAAI Conference on Web and Social Media*, pp. 774-785. Palo Alto, Calif.: The AAAI Press.

Zhou, E., Cao, Z. e Yin, Q. (2015). «Naive-Deep Face Recognition: Touching the Limit of LFW Benchmark or Not?».ArXiv:1501.04690.

Zhu, J., Luo, J., You, Q. e Smith, J.R. (2013). «Towards Understanding the Effectiveness of Election Related Images in Social Media». In *2013 IEEE 13th International Conference on Data Mining Workshops*, pp. 421-425. Piscataway, N.J.: IEEE.

Gaetano Scaduto
Università di Torino
Dipartimento di Culture, Politica e Società
Lungo Dora Siena, 100 – Torino
gaetano.scaduto@edu.unito.it
https://orcid.org/0000-0002-1368-2077